ΚΕΦΑΛΑΙΟ 1
Εισαγωγή

1.1 Η διττή σημασία της έννοιας «στατιστική»

Ίσως ο όρος «στατιστική» να σας φέρνει στο μυαλό αρχικά μεγάλες στήλες αριθμών ή εκφράσεις όπως:

«Οι στατιστικές λένε τα καλύτερα ψέματα»,
«Μην εμπιστεύεσαι καμία στατιστική, αν δεν την έχεις παραποιήσει πρώτα εσύ ο ίδιος».

Η έννοια «στατιστική» έχει πολλές σημασίες. Αφενός περιγράφει την απεικόνιση μέσω πίνακα ή τη γραφική απεικόνιση ενός συγκεκριμένου υπάρχοντος υλικού δεδομένων, όπως η στατιστική των τροχαίων ατυχημάτων για ένα συγκεκριμένο τμήμα της εθνικής οδού, η στατιστική των εισαγωγών μιας χώρας, διαιρεμένη κατά συγκεκριμένες ομάδες προϊόντων, χώρες προέλευσης κτλ. Αφετέρου η έννοια «στατιστική» περιγράφει το σύνολο των μεθόδων οι οποίες είναι ενδεδειγμένες για την απόκτηση και επεξεργασία εμπειρικών πληροφοριών.

Η στατιστική μεθοδολογία, η οποία από εδώ και στο εξής θα αναφέρεται εν συντομία ως στατιστική, αποτελεί το βασικό αντικείμενο του παρόντος εγχειριδίου. Απλές διαπιστώσεις σχετικά με δεδομένα που μετριούνται σε αριθμούς, π.χ. «Στις 27.05.1970 ο πληθυσμός που κατοικούσε στη Βαυαρία ανερχόταν σε 10.479.400 άτομα», δεν δημιουργούν προβλήματα και φυσικά δεν αποτελούν αντικείμενο της στατιστικής μεθοδολογίας. Αντίθετα, οι παραπάνω εκφράσεις συνδέονται περισσότερο με τη στατιστική μεθοδολογία, διότι μέσω αυτών επισημαίνεται εν τέλει πόσο προβληματική μπορεί να είναι η εξαγωγή συμπερασμάτων από στατιστικά δεδομένα. Θα επανέλθουμε όμως στη συνέχεια αρκετές φορές στο «στατιστικό ψέμα».[1]

1.2 Επιλογή της ύλης

Οι στατιστικές μέθοδοι μπορούν ουσιαστικά να χρησιμοποιηθούν σε όλες εκείνες τις περιπτώσεις στις οποίες συλλέγονται και αξιολογούνται εμπειρικά δεδομένα, ανεξάρτητα για ποιον επιστημονικό τομέα ή τομέα εφαρμογής πρόκειται. Ωστόσο η στατιστική χρησιμοποιείται ιδιαίτερα συχνά στους τομείς της οικονομίας των επιχειρήσεων, της οικονομικής επιστήμης, της ψυχολογίας, της κοινωνιολογίας, της παιδαγωγικής, της ιατρικής, της βιολογίας, της αγροτικής οικονομίας, της τεχνικής και της μετεωρολογίας.

1. Για πληροφοριακά παραδείγματα μπορείτε να ανατρέξετε κυρίως στον Krämer (2008) και στη βιβλιογραφία που αναφέρεται εκεί.

Ενώ η στατιστική σε αυτούς τους τομείς παίζει κατά κύριο λόγο τον ρόλο μιας χρήσιμης βοηθητικής επιστήμης, αποκτά σημαντικότερο ρόλο σε όλες τις νεο-αναπτυγμένες «-μετρίες», όπως π.χ. στη βιομετρία, τη δημομετρία, την οικονομετρία, την ψυχομετρία και την κοινωνιομετρία. Παρά αυτό το διευρυμένο φάσμα εφαρμογής εμείς θα περιοριστούμε κατά την επιλογή της ύλης, των παραδειγμάτων και των ασκήσεων στον τομέα των οικονομικών και κοινωνικών επιστημών.

Στο Πρώτο Μέρος εξετάζεται η **περιγραφική στατιστική**. Εδώ θα αναφερθούμε σε συνήθεις όρους που χρησιμοποιούνται από την πρακτική της οικονομικής και τον οικονομικό Τύπο όπως «αριθμοδείκτες, εποχική εκκαθάριση, προσδιορισμός τάσης, συσχέτιση» κτλ.

Στο Δεύτερο Μέρος παρουσιάζεται η **θεωρία πιθανοτήτων**, στον βαθμό που απαιτείται για την κατανόηση των ακόλουθων στατιστικών μεθόδων.

Το Τρίτο Μέρος εξετάζει εκείνες τις περιπτώσεις εφαρμογής, στις οποίες δεν μπορεί να πραγματοποιηθεί ή είναι αντιοικονομική μια πλήρης συλλογή δεδομένων. Έτσι, για παράδειγμα, δεν έχει νόημα να εξετάσει κανείς τη συνολική παραγωγή ενός μαζικού προϊόντος ως προς τη διάρκεια ζωής καθενός από τα προϊόντα αυτά, διότι τότε δεν θα περίσσευε κανένα κομμάτι προς πώληση. Παρομοίως, στις έρευνες αγοράς συχνά δεν μπορούν πάντα να ερωτηθούν όλοι οι δυνητικοί συμμετέχοντες στην αγορά, καθώς μια καθολική έρευνα θα ήταν υπερβολικά χρονοβόρα και δαπανηρή. Το ίδιο ισχύει φυσικά και για τα προβλήματα της δημοσκόπησης. Αντί για μια καθολική έρευνα διεξάγει κανείς μια μερική έρευνα, με την οποία επιλέγονται κάποια ερευνητικά υποκείμενα με βάση συγκεκριμένους νόμους της τύχης. Πραγματοποιεί δηλαδή κανείς μια δειγματοληπτική έρευνα. Το συμπέρασμα από τη δειγματοληπτική έρευνα για τον πληθυσμό είναι αποστολή της **επαγωγικής στατιστικής**. Οι κλασικοί επιμέρους τομείς της επαγωγικής στατιστικής, κυρίως η σημειακή εκτίμηση, η εκτίμηση διαστήματος και ο έλεγχος υποθέσεων, αποτελούν το περιεχόμενο του Τρίτου Μέρους.

Η συμπερίληψη του Τέταρτου Μέρους, το οποίο προσφέρει μια σύντομη **επισκόπηση σε περαιτέρω επιμέρους τομείς της στατιστικής**, οφείλεται στο γεγονός ότι μαζί με την επέκταση του εύρους εφαρμογής έχει πραγματοποιηθεί και μια σημαντική επέκταση του οπλοστασίου των στατιστικών μεθόδων. Καθώς ευνοείται από την αυξανόμενη δυνατότητα απόδοσης των υπολογιστικών συστημάτων επεξεργασίας δεδομένων (ΥΣΕΔ) και την αυξανόμενη ζήτηση για πληροφορίες, η οποία πρέπει να ικανοποιηθεί με το μικρότερο δυνατό κόστος, αυτή η τάση θα ισχύει ασφαλώς και για το μέλλον. Μερικοί τομείς, όπως ίσως ο υπολογισμός προγνώσεων, οι μέθοδοι πολλαπλών μεταβλητών, η στατιστική θεωρία λήψης αποφάσεων, το στατιστικό λογισμικό, για να ονομάσουμε ορισμένους μόνο, δεν μπορούν, παρά τη συχνή εφαρμογή τους, να συμπεριληφθούν ταυτόχρονα σε ένα βασικό μάθημα στατιστικής. Η μεταφορά τους σε μαθήματα βασικών σπουδών οδηγεί σύμφωνα με την εμπειρία μας στο γεγονός ότι μόνο ειδικοί μεταξύ των φοιτητών προσεγγίζουν τέτοιους τομείς. Η τάση που επικρατεί να χρησιμοποιεί κανείς στη μετέπειτα επαγγελματική του ζωή μόνο αυτό που γνωρίζει από τις σπουδές του ή τουλάχιστον αυτό που του είναι οικείο ως όνομα, συνηγορεί στο να παρουσιάσουμε εδώ κάποιους από αυτούς τους σημαντικούς μερικούς τομείς τουλάχιστον ως προς τα βασικά χαρακτηριστικά τους.

ΜΕΡΟΣ ΠΡΩΤΟ:
Περιγραφική στατιστική

Στις στατιστικές έρευνες εμφανίζονται κατά κανόνα χιλιάδες μεμονωμένα δεδομένα. Οι μέθοδοι της περιγραφικής στατιστικής (ή διερευνητικής στατιστικής) έχουν ως στόχο να χαρακτηρίσουν αυτή την ασαφή ποσότητα δεδομένων με όσο το δυνατόν λιγότερους –αλλά ουσιαστικούς– αριθμούς. Στην ακραία περίπτωση χρησιμοποιείται μόνο ένας και μοναδικός αριθμός για τον χαρακτηρισμό της συνολικής ποσότητας δεδομένων. Έτσι μια παράμετρος θέσης χαρακτηρίζει την κατάταξη μεγεθών όλων των μεμονωμένων δεδομένων, ένας δείκτης τιμών τη μεταβολή όλων των τιμών ενός εκτεταμένου καλαθιού προϊόντων, ένας συντελεστής συσχέτισης την εξάρτηση ανάμεσα στις εκφράσεις δύο διαφορετικών χαρακτηριστικών κ.ο.κ.

Φυσικά μια τόσο ευρύτατα επεκταμένη προετοιμασία και πύκνωση δεν μπορεί να πραγματοποιηθεί με έναν και μοναδικό τρόπο. Για τον λόγο αυτό, για την αιτιολογημένη επιλογή μιας κατάλληλης μεθόδου και για την αποτελεσματική ερμηνεία του αριθμητικού αποτελέσματος είναι απολύτως απαραίτητη μια θεμελιωμένη γνώση των βασικών προϋποθέσεων και ιδιοτήτων των διαφορετικών μεθόδων αξιολόγησης.

ΚΕΦΑΛΑΙΟ 2
Βασικές έννοιες των στατιστικών ερευνών

2.1 Χαρακτηριστικά, τιμές χαρακτηριστικών, φορείς χαρακτηριστικών, στατιστική μάζα

Σίγουρα κάποια στιγμή συμπληρώσατε και εσείς ένα ερωτηματολόγιο, και για να το κάνετε αυτό χρειάστηκε να δώσετε πληροφορίες σχετικά με την ηλικία, το φύλο, την οικογενειακή κατάσταση, τον τόπο κατοικίας, τη θέση εργασίας, το ύψος του εισοδήματος, το είδος της ασφάλισης υγείας, τον απαιτούμενο χρόνο για να φτάσετε στη δουλειά κ.ο.κ. Την κατανοητή δυσαρέσκεια που νιώσατε όταν το κάνατε δεν θα την αναλύσουμε περαιτέρω εδώ, αλλά θα διευκρινίσουμε μια σειρά από ειδικούς όρους. Στη στατιστική ορολογία λειτουργήσατε ως **φορέας χαρακτηριστικών** ή ως **στατιστική μονάδα,** στην οποία ερευνήθηκαν τα **χαρακτηριστικά στοιχεία** ηλικία, φύλο κτλ. **Εκφράσεις ή τιμές των χαρακτηριστικών** ονομάζονται τόσο οι διαφορετικοί αριθμοί τους οποίους μπορεί να λάβει ένα **ποσοτικό** χαρακτηριστικό (όπως η ηλικία, το ύψος του εισοδήματος), όσο και οι διαφορετικές κατηγορίες οι οποίες μπορούν να παρατηρηθούν σε ένα **ποιοτικό** χαρακτηριστικό (όπως το

φύλο, το θρήσκευμα). Οι φορείς χαρακτηριστικών που είναι σημαντικοί για μια στατιστική έρευνα συνοψίζονται σε ένα **βασικό σύνολο πληθυσμού**, στο οποίο μπορούμε να αναφερόμαστε απλά και με τον όρο «πληθυσμός». Εάν σε μια στατιστική έρευνα καταγραφεί πλήρως το βασικό σύνολο πληθυσμού, τότε μιλάμε για μια **πλήρη** ή **καθολική έρευνα**, ενώ σε άλλη περίπτωση μιλάμε για μια **μερική** ή **δειγματοληπτική έρευνα**.

Παράδειγμα 2.1:

- Η τελευταία απογραφή του πληθυσμού (ημέρα έρευνας 25.05.1987) αποτελεί μια καθολική έρευνα του πληθυσμού των κατοίκων της Ομοσπονδιακής Δημοκρατίας της Γερμανίας.
- Εάν σας παρουσιάστηκε από έναν υπάλληλο της Εθνικής Στατιστικής Υπηρεσίας το ερωτηματολόγιο που αναφέρθηκε αρχικά στο πλαίσιο της «αντιπροσωπευτικής στατιστικής του πληθυσμού και της επαγγελματικής ζωής» που χαρακτηρίζεται ως Mikrozensus (μικροαπογραφή), τότε ανήκετε σε ένα από τα 230.000 νοικοκυριά περίπου, τα οποία επιλέχθηκαν σύμφωνα με μια συγκεκριμένη διαδικασία δειγματοληπτικής έρευνας.
- Κατά την καταμέτρηση των χειροτεχνών του 1968 πραγματοποιήθηκε μια δειγματοληπτική έρευνα, η οποία περιλάμβανε περίπου 620.000 επιχειρήσεις που ήταν καταγεγραμμένες στο επάγγελμα του χειροτέχνη. Ως χαρακτηριστικά ερευνήθηκαν μεταξύ άλλων: η νομική μορφή, η ύπαρξη θυγατρικών, ο αριθμός των απασχολούμενων, τα ημερομίσθια, οι μισθοί, οι κοινωνικές παροχές, η χρήση πρώτων υλών, οι εξαγωγικές παραγωγικές σειρές.

Για την περιγραφική στατιστική –σε αντίθεση με την επαγωγική στατιστική– είναι ασήμαντο αν το προς επεξεργασία υλικό δεδομένων προέρχεται από μια συνολική ή από μια μερική έρευνα. Στη συνέχεια με τον όρο **στατιστική μάζα** θα εννοούμε είτε τον πληθυσμό ή ένα από τα μερικά σύνολα αυτού. Σε περίπτωση που εξετάζεται μόνο ένα μερικό σύνολο, τότε στο πλαίσιο της περιγραφικής στατιστικής οι διαπιστώσεις ισχύουν μόνο για αυτό το μερικό σύνολο.

2.2 Διάφοροι τύποι στατιστικής μάζας

Υπάρχουν στατιστικές μάζες τις οποίες συνδέει κανείς αυτονόητα με ένα συγκεκριμένο **χρονικό σημείο**. Αυτές οι στατιστικές μάζες χαρακτηρίζονται ως **αποθεματικές μάζες**· παραδείγματα είναι το απόθεμα μιας αποθήκης ή ο πληθυσμός μιας χώρας. Άλλες στατιστικές μάζες αναφέρονται αυτονόητα σε μια **χρονική περίοδο**. Αυτές οι στατιστικές μάζες χαρακτηρίζονται ως **μάζες κίνησης (ροής)** ή ως **μάζες γεγονότος (περιστατικού)**· κάποια παραδείγματα είναι τα αυτοκινητικά ατυχήματα ενός έτους ή η εκροή αποθεμάτων μιας εβδομάδας. Σε μια μάζα κίνησης οι φορείς χαρακτηριστικών έχουν μια αμελητέα μικρή «διάρκεια ζωής», ενώ αντίθετα σε μια μάζα αποθέματος όχι.

Λόγω κόστους, είναι πιο σύνηθες οι εκτεταμένες αποθεματικές μάζες να επικαιροποιούνται με την προσθήκη νέων δεδομένων, παρά να συλλέγονται από την αρχή δεδομένα σε μια νέα έρευνα. Κατά την ενημέρωση η στατιστική μάζα της προηγούμενης ημερομηνίας δειγματοληπτικής έρευνας αυξάνεται κατά τις **αντίστοιχες μάζες κίνησης**

των εισροών που έγιναν μεταξύ των δειγματοληπτικών ημερομηνιών και μειώνεται κατά τις αντίστοιχες μάζες κίνησης των εκροών.

Άσκηση 2.2: Πρόκειται για αποθεματικές μάζες ή για μάζες κίνησης;

α) Δοχεία ζύμωσης μιας ζυθοποιίας
β) Θανατηφόρα εργατικά δυστυχήματα σε μια περιοχή
γ) Νέες προσλήψεις μιας επιχείρησης
δ) Συνεργάτες μιας επιχείρησης
ε) Υποκαταστήματα μιας τράπεζας

2.3 Διαφορετικοί τύποι χαρακτηριστικών, κλιμάκωση, ταξινόμηση

Στην ενότητα 2.1 αναφέρθηκε ήδη ότι ως **ποσοτικά** ονομάζονται εκείνα τα χαρακτηριστικά των οποίων οι εκφράσεις, δηλαδή οι παρατηρήσιμες καταστάσεις ή τιμές τους, είναι αριθμοί· όλα τα άλλα χαρακτηριστικά θα λέγονται **ποιοτικά**. Οι εκφράσεις των ποιοτικών χαρακτηριστικών αποδίδονται κυρίως μέσω μιας λεκτικής διατύπωσης (όπως π.χ. καλά, μέτρια, αρνητικά, άγαμος, διαζευγμένος κτλ.). Εάν αντικαταστήσει κανείς αυτές τις λεκτικές διατυπώσεις με αριθμούς, εάν δηλαδή πραγματοποιήσει κανείς μια **ποσοτικοποίηση**, τότε ένα ποιοτικό χαρακτηριστικό γίνεται τυπικά ποσοτικό χαρακτηριστικό.

Παράδειγμα 2.3: Οι αποχρώσεις των ΙΧ αυτοκινήτων ποσοτικοποιούνται μέσω του αριθμού παραγωγής, π.χ. πράσινο = 117, μπλε = 410 κ.ο.κ. Σε μια στατιστική έρευνα το θρήσκευμα κωδικοποιείται π.χ. με τη μορφή: κανένα = 0, καθολικός = 1, ευαγγελιστής = 2, λοιπά = 3. Η νομική μορφή μιας επιχείρησης μπορεί να ποσοτικοποιηθεί π.χ. ως εξής: μετοχική εταιρεία = 1, ΕΠΕ = 2, εμπορική επιχείρηση = 3, ετερόρρυθμη εταιρεία = 4, ατομική επιχείρηση = 5, λοιπά = 6.

Η ποσοτικοποίηση, ωστόσο, εξαλείφει μόνο εξωτερικά τη διαφορά μεταξύ ενός (αυθεντικά) ποσοτικού και ενός (ποιοτικού αλλά) ποσοτικοποιημένου χαρακτηριστικού. Κατά τη στατιστική επεξεργασία του υλικού των δεδομένων πρέπει να δοθεί σημασία στο αν πραγματοποιήθηκε ποσοτικοποίηση ή όχι.

Παράδειγμα 2.4: Όταν δύο άτομα ανήκουν στο θρήσκευμα που σύμφωνα με το παράδειγμα 2.3 ποσοτικοποιείται ως θρήσκευμα 1 (=καθολικός) και 3 (=λοιπά), τότε δεν έχει νόημα να μιλάμε για το «μεσαίο θρήσκευμα» 2 (ευαγγελιστής). Εάν αντίθετα δύο άτομα αποκτούν 1000 ευρώ και 3000 ευρώ ως δώρο Χριστουγέννων, τότε μπορεί κανείς να μιλήσει για ένα «μέσο δώρο Χριστουγέννων» ύψους 2000 ευρώ.

Η πρακτική λογική η οποία κρύβεται πίσω από τις αριθμητικές εκφράσεις των χαρακτηριστικών, περιλαμβάνεται στη στατιστική μέσα από την έννοια της **κλιμάκωσης**. Παρακάτω θα εξηγήσουμε εν συντομία τις τρεις σημαντικότερες κλιμακώσεις.

Μια **ονομαστική κλίμακα** υπάρχει όταν οι εκφράσεις του εξεταζόμενου χαρακτηριστικού πρέπει μόνο να διαχωριστούν μέσω αριθμών που τους αποδίδονται. Όλα τα χαρακτηριστικά που παρουσιάζονται στο Παράδειγμα 2.3 έχουν ως βάση τους μια ονομαστική κλίμακα. Εφόσον οι αποδιδόμενοι αριθμοί έχουν μια καθαρά περιγραφική λειτουργία και μόνο, μπορούν να μετατραπούν αυθαίρετα σε άλλους αριθμούς μέσω μιας αντιστοιχίας.

Μια **ιεραρχική ή διατακτική** κλίμακα υπάρχει όταν οι εκφράσεις-τιμές του εξεταζόμενου χαρακτηριστικού δεν διαφοροποιούνται απλά αλλά μπορούν και να ταξινομηθούν σύμφωνα με μία ιεραρχική τάξη, η οποία μπορεί να αντικατοπτρίζεται μέσω των αποδιδόμενων σε αυτήν αριθμών. Οι ιεραρχικές κλίμακες αποτελούν τη βάση μέτρησης πολλών χαρακτηριστικών των κοινωνικών επιστημών, όπως της επιθετικότητας, της πίστης στην αυθεντία, της ευφυΐας, του κοινωνικού στάτους κ.ο.κ. Εφόσον στους αποδιδόμενους αριθμούς τίθεται μόνο η απαίτηση να αναπαράγουν τη σωστή ιεραρχική τάξη των τιμών του χαρακτηριστικού, αυτές μπορούν να μετασχηματιστούν αυθαίρετα σε άλλους αριθμούς μέσω ενός σχήματος που διατηρεί την ιεραρχική τάξη.

Μια **αριθμητική κλίμακα** ή μετρική κλίμακα υπάρχει όταν οι τιμές των εξεταζόμενων χαρακτηριστικών δεν μπορούν μόνο να μεταφερθούν σε μια ιεραρχική τάξη, αλλά επιπλέον μπορεί και να καθοριστεί σε τι βαθμό διαφοροποιούνται μεταξύ τους δύο διαφορετικές τιμές του χαρακτηριστικού.[1] Τυπικά παραδείγματα για τέτοιου είδους χαρακτηριστικά είναι όλα τα νομισματικά μεγέθη καθώς και οι μεταβλητές της φυσικής και της χημείας. Πολλές στατιστικές μέθοδοι πραγματοποιούνται μόνο σε χαρακτηριστικά αριθμητικής κλίμακας. Θα αναφερθούμε σε αυτά σε διάφορα σημεία του βιβλίου.

Εκτός από τη διάκριση των χαρακτηριστικών ως προς την κλιμάκωση, πρέπει επίσης να αναφέρουμε τον διαχωρισμό σε **διακριτά** και **συνεχή** χαρακτηριστικά. Ένα χαρακτηριστικό ονομάζεται **διακριτό**, όταν οι πιθανές εκφράσεις του διαμορφώνουν ποσότητα διακριτών αριθμών.[2] Τυπικά παραδείγματα είναι τα χαρακτηριστικά ονομαστικής κλίμακας και γενικά όλα τα χαρακτηριστικά που βασίζονται σε μια πράξη μέτρησης, όπως ο αριθμός των εργαζομένων μιας επιχείρησης, ο αριθμός των δωματίων ενός σπιτιού κτλ. Σε ένα **συνεχές χαρακτηριστικό** μπορούν αντίθετα για δύο εκφράσεις του να υλοποιηθούν και όλες οι ενδιάμεσες τιμές. Τυπικά παραδείγματα είναι εκείνα τα χαρακτηριστικά που βασίζονται σε μια όσο το δυνατόν πιο ακριβή σειρά μέτρησης, όπως συμβαίνει στη διάρκεια ζωής μιας βαλβίδας, τη διάμετρο ενός σωλήνα κτλ.

Για λόγους σκοπιμότητας ένα διακριτό χαρακτηριστικό, το οποίο μπορεί να λάβει πολλές τιμές, συχνά εξετάζεται ως συνεχές χαρακτηριστικό. Σε αυτή την περίπτωση χάριν ακριβείας γίνεται λόγος για ένα **οιονεί συνεχές** χαρακτηριστικό. Ακόμη και η αντίθετη διαδικασία, η εξέταση ενός συνεχούς χαρακτηριστικού ως διακριτού, μπορεί να ενδείκνυται για λόγους σκοπιμότητας. Για τον λόγο αυτό διαιρούμε –όπως στο Παράδειγμα 2.5– τον άξονα του χαρακτηριστικού σε τάξεις (= διαστήματα) και καταγράφουμε στη θέση της ακριβούς έκφρασης του χαρακτηριστικού μόνο την τάξη στην οποία εμπίπτει. Σε αυτή την περίπτωση μιλάμε για **ταξινόμηση**[3] ή για **ταξινομημένα δεδομένα,** σε ορισμένες δε περιπτώσεις και για **ομαδοποιημένα δεδομένα**.

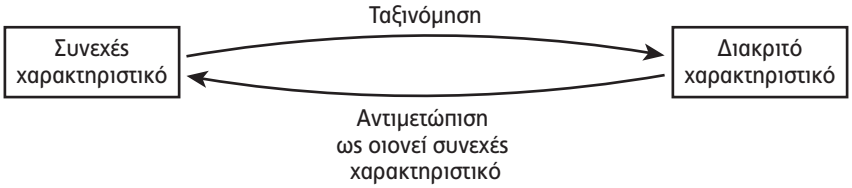
---

1. Με μεγαλύτερη ακρίβεια διαχωρίζει κανείς τη μετρήσιμη κλίμακα στην **κλίμακα διαστημάτων** (στην οποία οι αποστάσεις μεταξύ των τιμών μπορούν να συγκριθούν), την **κλίμακα σχέσεων** (στην οποία προστίθεται στην κλίμακα διαστημάτων ένα φυσικό μηδενικό σημείο) και την **απόλυτη κλίμακα** (στην οποία προστίθεται στην κλίμακα σχέσεων μια φυσική μονάδα). Για παράδειγμα, η θερμοκρασία μετριέται σε μια κλίμακα διαστημάτων (αν δεν μετριέται σε βαθμούς Κελσίου). Οι τιμές, τα μήκη και τα βάρη βασίζονται σε μια κλίμακα σχέσεων. Οι μοναδιαίοι αριθμοί βασίζονται σε μια απόλυτη κλίμακα.

2. Σ.τ.Ε.: Πεπερασμένο ή αριθμήσιμο σύνολο τιμών.

3. Αυτή μπορεί φυσικά να πραγματοποιηθεί και σε οιονεί συνεχή ή σε διακριτά χαρακτηριστικά.

Σχήμα 1: Μετατροπή ενός συνεχούς χαρακτηριστικού σε διακριτό χαρακτηριστικό και αντίστροφα



Παράδειγμα 2.5: Στο ερωτηματολόγιο για το Mikrozensus αναγράφονται (από το 1972) 15 κατηγορίες μεγέθους εισοδήματος. Αυτό οφείλεται στο γεγονός ότι οι συνεντευξιαζόμενοι πρώτον δεν γνωρίζουν συχνά με ακρίβεια το εισόδημά τους και δεύτερον είναι πολύ πιο πρόθυμοι να καταχωρίσουν τον εαυτό τους σε κάποια δεδομένη κατηγορία μεγέθους εισοδήματος από το να αναφέρουν με ακρίβεια τα διάφορα έσοδά τους. Ακόμη και στη Στατιστική Επετηρίδα επιλέγονται κατηγορίες τάξεων, όπως ο πίνακας που παρουσιάζεται στο σχήμα 2.

Σχήμα 2: Αποδόσεις που επηρεάζουν το εισόδημα και επιδόματα αποταμίευσης των εργαζομένων για το έτος 1974 (σε εκατομμύρια γερμανικά μάρκα) σύμφωνα με τον τρίτο νόμο περί σχηματισμού περιουσίας (από τη Στατιστική Επετηρίδα 1978 για την Ομοσπονδιακή Δημοκρατία της Γερμανίας, σ. 428)

Ακαθάριστη αμοιβή	**Φορολογούμενοι με αποδόσεις με επίδραση στο εισόδημα συνολικά**		
	Περιπτώσεις (σε 1000)	**Αποδόσεις με επίδραση στην περιουσία**	**Επιδόματα αμοιβής των εργαζομένων**
κάτω από 2400	146,0	37	11
2400 έως κάτω από 4800	378,7	158	47
4800 έως κάτω από 7200	424,4	194	58
7200 έως κάτω από 9600	366,7	165	50
9600 έως κάτω από 12000	468,9	223	67
12000 έως κάτω από 16000	1096,4	555	169
16000 έως κάτω από 20000	1791,7	960	296
20000 έως κάτω από 25000	2774,1	1518	473
25000 έως κάτω από 36000	4965,7	2749	821
36000 έως κάτω από 50000	3508,0	1986	592
50000 έως κάτω από 75000	1090,7	645	171
75000 έως κάτω από 100000	102,7	59	14
100000 και άνω	19,5	11	2
Σύνολο	17133,5	9259	2770

Άσκηση 2.6: Ποια από τα παρακάτω χαρακτηριστικά είναι συνεχή και ποια διακριτά;

α) Επάγγελμα
β) Φρενάρισμα ενός ΙΧ
γ) Εθνικότητα

Άσκηση 2.7: Σε ποια κλιμάκωση βασίζονται τα ακόλουθα χαρακτηριστικά;

α) Διάρκεια παραγωγής
β) Βαθμός σχολείου
γ) Βαθμός δυσκολίας μιας διαδρομής ανάβασης
δ) Διάμετρος κεφαλιού ενός ατόμου

2.4 Διαφορετικά είδη στατιστικών ερευνών

Πέραν της διαίρεσης των στατιστικών ερευνών σε συνολικές και μερικές έρευνες που αναφέρθηκε ήδη στην Ενότητα 2.1, υπάρχει μια σειρά περαιτέρω δυνατοτήτων διαχωρισμού που πρέπει να αναφερθούν. Σύμφωνα με το είδος της απόκτησης του υλικού δεδομένων μπορεί να γίνει διαχωρισμός ανάμεσα σε **έρευνα ερωτήσεων** (ταχυδρομική/διαδικτυακή μέσω ερωτηματολογίου ή προσωπικά μέσω συνέντευξης), **παρατήρηση** (μέτρηση κυκλοφορίας, μέτρηση του χρόνου αναμονής των πελατών πριν από το ταμείο ενός σουπερμάρκετ) και **πείραμα** (καταγραφή της συμπεριφοράς εξεταζομένων σε υποθετικές καταστάσεις λήψης αποφάσεων, μέτρηση της διάρκειας του ύπνου μέσω χρήσης διαφορετικών βοηθητικών μέσων ύπνου) καθώς και κατά διαφορετικό τρόπο ανάμεσα σε:

- πρωτογενή έρευνα, στην οποία το υλικό των δεδομένων συλλέγεται μόνο για την προγραμματισμένη έρευνα, και σε
- δευτερογενή έρευνα, κατά την οποία χρησιμοποιείται ήδη διαθέσιμο (και συλλεγμένο για άλλους σκοπούς) υλικό δεδομένων, όπως συμβαίνει με τη χρήση των φορολογικών καρτών για την έρευνα κατανομής του εισοδήματος.

Για τη διεξαγωγή μερικών ερευνών πρέπει να ονομάσουμε, εκτός από την αυθαίρετη επιλογή φορέων χαρακτηριστικών, η οποία δυστυχώς χρησιμοποιείται ακόμη στην πράξη, κυρίως τις διαφορετικές **τεχνικές δειγματοληπτικής έρευνας** οι οποίες αναφέρονται αναλυτικότερα στο Κεφάλαιο 18, καθώς και τις μεθόδους της **προγραμματισμένης επιλογής.** Οι δύο γνωστότερες μέθοδοι της προγραμματισμένης επιλογής είναι η διαδικασία αποκοπής και η επιλογή ποσόστωσης.

Κατά τη **διαδικασία αποκοπής,** την οποία χρησιμοποιεί συχνά η δημόσια στατιστική, περιλαμβάνονται μόνο οι φορείς χαρακτηριστικών με τις μεγαλύτερες συνεισφορές στα εξεταζόμενα χαρακτηριστικά. Έτσι λειτουργεί η Ομοσπονδιακή Στατιστική Υπηρεσία, για παράδειγμα στη βραχυπρόθεσμη έκθεση των παραγωγικών επιχειρήσεων μόνο οι επιχειρήσεις με έναν συγκεκριμένο κατώτερο αριθμό απασχολούμενων περιλαμβάνονται στη μηνιαία έρευνα.

Κατά την **επιλογή ποσόστωσης,** η οποία χρησιμοποιείται σπανίως από τη δημόσια στατιστική και συχνότερα από τα ινστιτούτα έρευνας αγοράς και έρευνας κοινής γνώμης, δίδεται μεγάλη σημασία στο ότι συγκεκριμένες εκφράσεις χαρακτηριστικών στο επιλεγμένο μερικό σύνολο πληθυσμού έχουν την ίδια σχετική συχνότητα όπως στο βασικό σύνολο. Αυτές οι (απόλυτες ή) σχετικές συχνότητες χαρακτηρίζονται ως ποσοστώσεις και δίδονται στους υπεύθυνους που παίρνουν τη συνέντευξη. Στο πλαίσιο των ποσοστώσεων, οι οποίες πρέπει να ορίσουν την «αντιπροσωπευτική εγκάρσια τομή του

πληθυσμού», ο αρμόδιος της συνέντευξης μπορεί να επιλέξει αυθαίρετα τους φορείς χαρακτηριστικών που θα ερωτηθούν.

Παράδειγμα 2.8: Το Ινστιτούτο Δημοσκόπησης του Allensbach χρησιμοποιεί τα ακόλουθα χαρακτηριστικά (σε παρένθεση οι ποσοστώσεις των εκάστοτε εκφράσεων) ως βάση για την επιλογή ποσοστώσεων από τον πληθυσμό ηλικίας άνω των 16 ετών:[4]

- Φύλο (άνδρες 47%, γυναίκες 53%)
- Ηλικία (16 έως 29 ετών 25%, 30 έως 44 ετών 28%, 45 έως 59 ετών 21%, 60 ετών και άνω 26%)
- Επαγγελματικές κατηγορίες (εργάτες και αγρότες 43%, εστιάτορες 5%, υπάλληλοι 34%, δημόσιοι υπάλληλοι 9%, ανεξάρτητοι σε τομείς εμπορίου και επιχειρήσεων 8%, ελεύθεροι επαγγελματίες 1%).
- Μεγέθη τόπου κατοικίας (κάτω από 2000 κατοίκους 8%, 2000 έως κάτω από 20000 κατοίκους 33%, 20000 έως κάτω από 100000 κατοίκους 24%, 100000 κάτοικοι και άνω 35%).
- Τοποθεσία της κατοικίας (Σλέσβιγκ-Χόλσταϊν 4%, Αμβούργο 3%, Βρέμη 1%, Κάτω Σαξονία 12%, Βόρεια Ρηνανία – Βεστφαλία 28%, Έσεν 9%, Ρηνανία-Πφαλτς 6%, Σάαρλαντ 2%, κ.ο.κ.).

Όταν σύμφωνα με τα παραπάνω πρέπει να ερωτηθούν 1000 άτομα, πρέπει ανάμεσά τους να βρίσκονται 470 άνδρες και 530 γυναίκες, επιπλέον 250 άτομα ηλικίας από 16 έως 29 ετών, 50 αγρότες κ.ο.κ.

2.5 Κριτική περίληψη, βιβλιογραφία

Στα ακόλουθα κεφάλαια δεν θα προβληματιστούμε περισσότερο για το διαθέσιμο υλικό δεδομένων, αλλά θα εξετάσουμε μόνο τις μεθόδους αξιολόγησης που χρησιμοποιούνται σε αυτό. Σε αυτό το σημείο πρέπει να κάνουμε κάποια κριτικά σχόλια σχετικά με τη δημιουργία του υλικού δεδομένων.

- Ο καθορισμός των εξεταζόμενων χαρακτηριστικών αποτελεί ήδη μια αποφασιστική τοποθέτηση. Εάν ο κατάλογος των εξεταζόμενων χαρακτηριστικών με την έννοια της σκόπιμης έρευνας είναι άσκοπος ή ελλιπής, τότε αυτό δεν μπορεί να διορθωθεί ακόμη και με τις πιο εκλεπτυσμένες μεθόδους αξιολόγησης. Εάν, για παράδειγμα, θέλει κανείς να αποκτήσει πληροφορίες σχετικά με το ποιοι παράγοντες είναι αποφασιστικοί για την άνοδο ενός εργαζομένου σε μια διευθυντική θέση, τότε είναι σκόπιμο να επικεντρωθεί σε χαρακτηριστικά όπως η εκπαίδευση, το ποσοστό ευφυΐας, η ικανότητα επιβολής κτλ. Ωστόσο δεν υπάρχει καμία εγγύηση ότι το πιο σημαντικό χαρακτηριστικό έχει συμπεριληφθεί στην έρευνα· θα μπορούσε π.χ. – το παράδειγμα είναι πλασματικό– να είναι και το μη υπολογισμένο χρώμα των

4. Πρβλ. Allensbacher Jahrbuch der Demoskopie 1976-1977, σ. 50. Εδώ λαμβάνονται υπόψη για τη συμπλήρωση του αντιπροσωπευτικού δείγματος του πληθυσμού τα επιπλέον χαρακτηριστικά οικογενειακή κατάσταση, θρήσκευμα και σχολική εκπαίδευση. Επειδή οι ανάλογες εκφράσεις χαρακτηριστικού των ατόμων δεν μπορούν να εξεταστούν «με μία ματιά», δεν δίδονται στους υπεύθυνους που παίρνουν τη συνέντευξη ποσοστώσεις.

ματιών. Δυστυχώς δεν υπάρχει καμία συνταγή ευρεσιτεχνίας για τη δημιουργία πλήρους καταλόγου χαρακτηριστικών. Η πιο κοντινή λύση της συμπερίληψης όσο το δυνατόν περισσότερων χαρακτηριστικών τις περισσότερες φορές δεν είναι πρακτική για λόγους κόστους και λόγω του υπερβολικού φόρτου εργασίας των ατόμων που παίρνουν τη συνέντευξη.

- Παρόμοια επιρροές νόθευσης μπορούν να προκύψουν από τη συνειδητή άρνηση απάντησης, ασαφείς διατυπώσεις στο ερωτηματολόγιο ή από την επιρροή που ασκεί το άτομο που παίρνει τη συνέντευξη. Η εμπειρική κοινωνική έρευνα αφιερώνει δικαίως ιδιαίτερη προσοχή σε αυτά τα προβλήματα. Για περισσότερες πληροφορίες βλ. Holm (1975) ή Atteslander (2008).
- Περαιτέρω επιρροές νόθευσης μπορούν να προκύψουν από την επιλογή «λανθασμένων» μερικών συνόλων. Ως ένα χαρακτηριστικό παράδειγμα αναφέρεται ένας υπεύθυνος συνέντευξης, ο οποίος πρέπει να διερευνήσει εμπειρικά τη γνώμη του συνολικού πληθυσμού στο θέμα «πληρωμή τελών για τις σπουδές» και για τον σκοπό αυτό επιλέγει αποκλειστικά φοιτητές, διότι θεωρεί ότι αυτοί είναι ιδιαίτερα πρόθυμοι να προσφέρουν πληροφορίες και επειδή τους συναντά σε μεγάλους αριθμούς στο εστιατόριο των φοιτητών. Περαιτέρω παραδείγματα βρίσκονται στον Krämer (2008).
- Νοθεύσεις μπορούν επίσης να ανακύψουν λόγω του σχηματισμού τάξεων, ιδιαίτερα μέσω της διαμόρφωσης ανοιχτών τάξεων άκρων.
- Επιπλέον η μεταβλητότητα της απάντησης είναι και αυτή ένα πρόβλημα που δεν πρέπει να παραβλέπεται. Σχετικά με αυτό, πρβλ. Strecker et al. (1983).

Ως επιπλέον συμπληρωτική βιβλιογραφία για όποιον θέλει να εντρυφήσει σχετικά με το Κεφάλαιο 2 προτείνονται: Ferschl (1985), Krug et al. (2001), Menges/Skala (1973), Pfanzagl (1983), Schwarze (2005), Wetzel (1971).